

日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日

Date of Application:

2001年 3月 2日

出 願 番 号

Application Number:

特願2001-057631

出 願 人

Applicant (s):

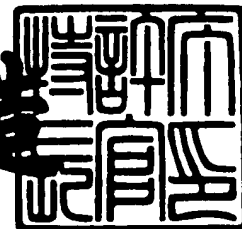
株式会社日立製作所

BEST AVAILABLE COPY

2001年 4月13日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造





IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

500.39944X00

Applicant(s): Y. MATSUMOTO, ET AL.
Serial No.: 09 / 820,947
Filed: MARCH 30, 2001
Title: "STORAGE SUBSYSTEM THAT CONNECTS FIBRE CHANNEL
AND SUPPORTS ONLINE BACKUP".

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

MAY 9, 2001

Sir:

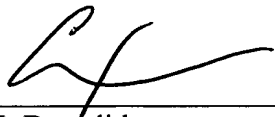
Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s)
the right of priority based on:

Japanese Patent Application No. 2001 - 057631
Filed: MARCH 2, 2001

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/rp
Attachment

【書類名】 特許願

【整理番号】 K00012211A

【提出日】 平成13年 3月 2日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

【氏名】 松本 佳子

【発明者】

【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

【氏名】 ▲高▼本 賢一

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 バックアップ処理可能な記憶システム

【特許請求の範囲】

【請求項 1】

上位装置からのデータを受け付けるディスク制御装置と、前記ディスク制御装置が受け取った前記データを記憶する複数の記憶装置からなる記憶システムにおいて、前記ディスク制御装置は、別の記憶システムと接続し前記記憶装置に記憶されている前記データを前記別の記憶システムにデータ転送することでデータのバックアップを行う記憶システムであって、

前記ディスク制御装置は、前記上位装置と前記別の記憶システムとを接続する一つのポート制御部を有し、前記ポート制御部は前記上位装置と前記記憶装置とのデータ転送と、前記別の記憶システムと前記記憶装置とのデータ転送を同時に行うことのできる 1 つのプロセッサを有していることを特徴とする記憶システム。

【請求項 2】

前記ディスク制御装置は、上位装置または他の記憶システムの少なくとも 1 つと接続する前記ポート制御部を複数有していることを特徴とする請求項 1 記載の記憶システム。

【請求項 3】

前記ディスク制御装置は、更に前記複数のポート制御部と前記複数の記憶装置と接続するディスクアレイ制御部を有し、

前記ディスクアレイ制御部は、前記複数のポート制御部の負荷率を保持するメモリと、前記メモリに記憶された前記負荷率に応じて、バックアップ処理を行うポート制御部を選択し実行させるプロセッサとを有していることを特徴とする請求項 2 記載の記憶システム。

【請求項 4】

前記ディスクアレイ制御部のプロセッサは、更に各ポート制御部毎に割り付ける処理種別情報を前記メモリに記憶し、前記処理種別情報に基づいて各ポート制御部の処理を実行させることを特徴とする、請求項 3 記載の記憶システム。

【請求項 5】

前記処理種別情報は、オンライン専用処理、バックアップ専用処理、またはオンライン・バックアップ同時処理のうちのいずれかの値であることを特徴とする請求項 4 記載の記憶システム。

【請求項 6】

前記ディスク制御装置は、前記ディスクアレイ制御部と前記複数のポート制御部とを有したコントローラを複数具備し、各コントローラに具備する前記ディスクアレイ制御部が相互に接続されていることを特徴とする請求項 3 記載の記憶システム。

【請求項 7】

前記ディスクアレイ制御部のプロセッサは、自コントローラ内に具備する前記メモリに記憶されている負荷率と、前記他のコントローラに具備する前記メモリに記憶されている負荷率とを参照し、バックアップ処理を実行すべきポート制御部を選択し実行させることを特徴とする請求項 6 記載の記憶システム。

【請求項 8】

前記コントローラは、更に自コントローラで発生した障害を検知する障害制御部を有し、前記障害制御部は前記ディスクアレイ制御部と他のコントローラ内の前記障害制御部とに接続し、前記他のコントローラ内の前記障害制御部より伝達された障害検知情報を前記ディスクアレイ制御部に伝達し、前記ディスクアレイ制御部は、前記他のコントローラで実行していたバックアップ処理を自コントローラに引き継ぐことを特徴とする請求項 6 記載の記憶システム。

【請求項 9】

前記ディスクアレイ制御部のプロセッサは、自コントローラ内に具備する前記メモリに記憶されている負荷率と前記他のコントローラに具備する前記メモリに記憶されている負荷率とを参照して前記他のコントローラで実行していたバックアップ処理を自コントローラに引き継ぐことを特徴とする請求項 6 記載の記憶システム。

【請求項 10】

前記ディスクアレイ制御部のプロセッサは、自コントローラ内の複数のポート制御部に、バックアップ処理を同時に処理させることを特徴とする請求項 6 記載の

記憶システム。

【請求項 1 1】

前記ディスク制御装置は、更にユーザインターフェースを備えた I / F 装置を具備し、各コントローラにおけるバックアップ処理の起動を指示することを特徴とする請求項 6 記載の記憶システム。

【請求項 1 2】

前記 I / F 装置は、前記各ポート制御部の処理種別情報を設定することを特徴とする請求項 1 1 記載の記憶システム。

【請求項 1 3】

前記ポート制御部は、前記上位装置および前記他の記憶システムとの間をファイバチャネルで接続されたことを特徴とする、請求項 1 ないし 1 2 に記載の記憶システム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、記憶システムに関し、特に高可用性を求められるコンピュータシステムに接続される。

【0 0 0 2】

【従来技術】

今般の計算機システムは、電子商取引との融合により 365 日 24 時間稼動といった長時間稼動が必要となり、また保持する情報量も年々増加しているため、データの保全策も検討する必要があった。その中で特に装置障害等によるデータ喪失があってもすぐに復旧できるよう、データのバックアップは重要な処理として位置付けられている。

【0 0 0 3】

一般にハードディスク等のコンピュータの外部記憶装置（記憶システム）に記録されたデータは、装置の障害、ソフトウェアの欠陥、誤操作等によりデータを喪失した場合に、喪失データを回復できる様に定期的にテープなどにコピーして保存（バックアップ）する。このとき、バックアップデータは、採取した時点での

データ記憶イメージで保存させる必要があり、かつバックアップ処理を実行しながら、通常のオンライン処理も実行させなければならない。そのために、オンライン処理でアクセスしているデータをデュプレックスし、バックアップ処理はデュプレックスされた一方のボリュームに対して実施し、バックアップ処理実行中のオンライン処理は、他方のボリュームに対して更新データを保持させながら行うことで実現されている。

【 0 0 0 4 】

従来、このようなバックアップの管理はホストコンピュータ側のプログラムにより行われていた。

【 0 0 0 5 】

例えば、米国特許 5,649,152 号公報に示されている技術では、あるデータの 2 重化の管理、任意時点のデータの保存、データその他記憶装置へのバックアップ転送指示等、全てホストコンピュータ内のバックアップ管理プログラムにて行われていた。

【 0 0 0 6 】

また、米国特許 5,051,887 号公報では、ホストコンピュータの管理により常に制御装置側にデータを二重に書き込んだ技術が示されている。この方式では、ホスト-制御装置間のデータ転送経路等に多大な負荷がかかっていた。

【 0 0 0 7 】

また、外部記憶装置によるデータの二重化による方法としては、例えば、コンピュータ内のプログラムが行っていたバックアップ管理を外部記憶装置内で行う技術がある。この例として、米国特許 5,845,295 号公報に示されている。これに示される技術では、二重化の管理、データの保存は外部記憶装置内で実現可能だが、実際にバックアップを行う時はホストコンピュータが外部記憶装置に対して READ 要求（以下 RD と記す）を行い、データを吸い上げ、他の記憶装置に対し、当該データの WRITE 要求（以下 WR と記す）を行うことにより実現していた。つまり、外部記憶装置としては、ホストからの指示に従い、ホストに対し、データを転送しているにすぎなかった。

【 0 0 0 8 】

その後ホストからの指示ではなく、外部記憶装置が直接他の記憶装置に対してデータコピーを実現する方法が米国特許 5, 7 4 2, 7 9 2 号公報において提案された。これは、オンライン処理を実施しながらバックグラウンドでデータのコピーを行うものである。この技術はバックアップとしても利用可能であり、これにより長時間連続稼動するなかでデータのバックアップ処理を実行できるようになった。

【 0 0 0 9 】

【発明が解決しようとする課題】

従来技術で示したとおり、長時間連続稼動のなかでのデータバックアップ処理を同時に連続稼動させる技術は確立しているが、これらの技術で開示されている外部記憶装置は主に大型外部記憶装置に適した技術である。データバックアップ処理とオンライン処理を実施する際、大型外部記憶装置のデータ転送口、すなわち、ホストとの接続口と、バックアップ装置との接続口（以下ポートとよぶ）を別々のハードウェアで実施することで実現されている。これは、大型外部記憶装置が冗長度を持たせるために複数のハードウェアを実装しているからである。こうした大型外部記憶装置の場合に、従来技術は適用可能である。

【 0 0 1 0 】

ところが、ワークステーションやパーソナルコンピュータなどに接続された小型外部記憶装置の場合には、冗長度を持たせることでコストがかかり、小型としてのメリットが活かせない問題がある。小型外部記憶装置も、大型外部記憶装置と同様、長時間連続稼動中の同時バックアップ処理を実現する要求が高まっており、如何に少ないハードウェアで実現するかが第1の課題であった。

【 0 0 1 1 】

また、従来のホスト指示によるバックアップ方式では、二重化処理中にデータ更新を行うと、外部記憶装置とホストとの転送が4回（二重化により2回分WR（Write）、バックアップ処理にてRD（Read）、そして他記憶装置へのWR（Write））発生することになり、ホストー記憶装置間のバスの負荷が増大し、通常のアプリケーションのパフォーマンスにまで影響せざるを得なかった。

【 0 0 1 2 】

特に、大量のデータをバックアップする際には、外部記憶装置内のプロセッサ処理能力がバックアップ処理に割かれるため、バックアップ終了までの長時間にわたり性能が低下する。最近、外部記憶装置内での二重化の機能が普及しているが、当該機能だけでは、バックアップの際の通常のオンライン処理に対するパフォーマンスの低下は防止できない。また、オープン向けの中小規模ディスクアレイ制御装置等では、コントローラとホストとのデータ転送経路は1つしか存在せず、この1つの経路がバスであった場合、バックアップを実現するためには、通常のオンライン処理を停止させるしかなかった。

【0013】

本発明の第一の目的は、ホストとバックアップ用外部記憶装置との接続口であるポートを1つにし、かつ1つのポートでバックアップ処理と通常のオンライン処理とを同時にまたは切り替えながら実行することのできるポート制御部を有した外部記憶装置を実現することにある。

【0014】

本発明の第2の目的は、オンライン処理のパフォーマンスを劣化させることなくバックアップ処理を実行可能な記憶制御装置を提供することにある。

【0015】

【課題を解決するための手段】

上記課題を解決する為に、記憶システムにおいて、ホスト及び他記憶制御装置とのデータ転送を行う単一の通信ポート、制御情報とホストからのデータを格納するキャッシュメモリ、ホストからのデータを格納するディスク装置群等の記憶装置、通信ポートとキャッシュメモリと記憶装置を制御する制御部を設け、通信ポートを有するポート制御部において通信処理を多重化させることで実現する。ホストと外部記憶装置およびバックアップ用の他外部記憶装置との間を F i b r e N e t w o r k で接続する。制御情報は、通信ポート毎のホストからの I/O 要求を実行する為の情報、他の記憶制御装置へのバックアップを行う為の転送情報、通信ポート毎の負荷情報、バックアップの進捗状態を管理する情報、ユーザーがポート毎にオンライン処理専用・バックアップ専用・オンライン/バックアップ共用を選択するための情報、データ二重化を制御するための情報、二重化

とバックアップのタイミングをユーザーが指示するための情報等からなり、これらは必要に応じてキャッシュメモリに記録され使用される。

【0016】

【発明の実施の形態】

以下、本発明の1実施例を図面を用いて説明する。

【0017】

図1は、本実施例に関するディスクアレイ装置を含むシステム全体の1構成例である。図1に於いて、10、20、30はホストコンピュータ（上位装置）であり、5000、6000はディスクアレイ装置（上位装置に対しての外部記憶装置（記憶システム））であり、各々のホスト及びディスクアレイ装置はFabric SwitchによりFibre Channelで接続され、ストレージエリアネットワーク（SAN）40を構成する。ディスクアレイ装置5000、6000は、ホストからのオンライン処理を各々実行してもいいし、ディスクアレイ装置6000は、ディスクアレイ装置5000のバックアップ用装置としてもよい。

【0018】

本実施例ではディスクアレイ制御装置5000がホストからのオンライン処理を実行しながら、データをデュプレックスし、ディスクアレイ制御装置6000にバックアップデータを転送するものとする。尚、バックアップデータを受領する外部記憶装置（記憶システム）は、ディスクアレイ装置以外でもかまわない。例えば、バックアップ装置は、磁気テープライブラリ、及び光ディスク装置でもよい。

【0019】

図2は、ディスクアレイ装置の構成を示す。ディスクアレイ装置5000は、デュアル構成をとるコントローラA部1000、コントローラB部2000、ホストからのデータを格納するディスク装置群4000、コントローラ内のポート制御部の動作の指定等を行うPC/WS（3000）から構成される。ディスク装置群4000は、アレイ構成であることが多い。

【0020】

次に、コントローラ A、B について説明する。コントローラ A、B とも同じ構成であり、コントローラ A について以下説明する。

【0021】

コントローラ A は、それぞれ、キャッシュメモリ 500 と、データ転送を行うポート A (101) および当該ポート A を制御するポート A 制御部 100 と、データ転送を行うポート B (201) および当該ポート B を制御するポート B 制御部 200 とを備える。これらポートは、Fibre Channel に接続され Fabric Switch によりネットワークを構成している。

【0022】

更に、コントローラ A は、コントローラ A 全体を制御するディスクアレイ制御部 300 を有し、さらにディスクアレイ制御部 300 は、当該コントローラを制御するための各種情報を格納する RAM 400 を有する。また、ディスクアレイ制御部 300 は、タイマー機能を併せ持っている (図 2 には図示しない)。

【0023】

なお、前述のキャッシュメモリ 500 は不揮発メモリであることが多い。また、本実施例では、キャッシュメモリ 500 内のデータは、コントローラ B のキャッシュメモリに 2 重書きされる。キャッシュメモリ 500 はホストからのデータを一時的に貯えるデータ部以外にデータの情報、状態を管理する管理情報部 600 を持つ。

【0024】

更に、コントローラ A は、ディスク装置群 4000 を制御するディスク装置制御部 (ドライブ制御部) 700、800 を有する。ディスク装置制御部は、ディスク装置群内のディスク装置とのデータ転送を行う。ディスク装置は、データを記憶する記憶ドライブとこれを制御するドライブ制御部からなり、SCSI インターフェース、又は Fibre Channel にてディスク制御装置と接続される。

【0025】

本実施例では、コントローラ A は上位装置 (ホストコンピュータ) と Fibre Channel を経由して接続されており、コントローラ 当たり 2 つのポー

トを持つ。障害制御部 9 0 0 はコントローラ A 内の回復不可能な障害を検出すると、コントローラ B に検出した障害を通知する機能を持つ。報告を受けた、コントローラ B は、キャッシュメモリ内のデータ・管理情報を引き継ぎ、コントローラ A で行われていた処理を引き継ぐ。

【 0 0 2 6 】

図 3 にディスク装置群の構成を示す。ディスク装置群はアレイ構成をとり、複数の論理ボリュームを割り当てることができる。また、バックアップ用に割り当てられる論理ユニットの副ボリュームは論理ボリューム(正)毎に設定されていて、もしいし、あるボリューム群がワークとして割り当てられていて、その部分を 2 重化する時に使用してもよい。

【 0 0 2 7 】

図 4 にポート制御部 1 0 0、2 0 0 とディスクアレイ制御部 3 0 0 間の I / F (インタフェース) 情報を示す。I / F 情報は、ポート毎に他記憶制御装置への転送を指示する転送指示情報 4 1 0 (ポート A 用)、4 3 0 (ポート B 用) と、ホストからの I / O 要求をディスクアレイ制御部 3 0 0 に伝えるホスト要求情報 4 2 0 (ポート A 用)、4 4 0 (ポート B 用) から構成される。各情報は、ポート ID (4 1 1)、コマンド 4 1 2、キャッシュ ADR (ADDRESS) 情報 4 1 3 から構成される。ポート ID は、転送指示情報の場合はストレージエリアネットワーク上のどの装置に対するデータ転送要求かを識別する情報であり、ホスト要求情報の場合はどの装置からのデータ転送要求なのかを識別する情報である。ポート ID は、具体的には、Fibre Channel の AL-PA (Fibre Channel Arbitrated Loop or FC-AL) に相当する物である。また、コマンド情報 4 1 2 は、RD / WR 等を識別する情報であり、キャッシュ ADR 情報 4 1 3 は、記憶装置内での当該データの格納位置、又は転送位置を指示するから構成される。これら情報は、RAM 内で管理される。

【 0 0 2 8 】

ポート A ホスト要求情報 4 2 0 は、ポート A 制御部が、Fibre Channel にて受領したホストからの I / O 要求を受信したとき、ディスクアレイ制御部 3 0 0 に I / O 要求を伝えるために設定される。ポート B ホスト要求情報 4

4 0 は、ポート A ホスト要求情報 4 2 0 と同様に、ポート B 制御部によって設定される。ディスクアレイ制御部は 4 2 0、4 4 0 の情報を参照し、ディスク装置 4 0 0、キャッシュメモリ 5 0 0 を制御し、当該 I/O 要求を実行する。

【 0 0 2 9 】

一方、ポート A 転送指示情報 4 1 0 は、ディスク装置群 4 0 0 の論理ボリューム（副）からのバックアップデータを他記憶制御装置に転送する時に設定される。ポート A 制御部 1 0 0 は、本情報 4 1 0 を参照し、Fibre Channel を介してコマンドを発行する。他記憶装置へコマンドを発行する場合は、ポート ID に、他記憶装置のポート ID を指定する値が設定されることになる。他記憶制御装置は本コマンドを受領し、実行することにより、バックアップデータを記憶媒体に格納することができる。ポート B 転送指示情報 4 2 0 も、ポート A 転送指示情報 4 1 0 の場合と同様に、ポート B 制御部 2 0 0 により使用される。

【 0 0 3 0 】

ポート A 転送指示情報 4 1 0、およびポート A ホスト要求情報 4 2 0 が、ホストおよびバックアップ装置とのデータ通信において、ポート制御部 1 0 0 とディスクアレイ制御部との I/F で利用される点について説明した。この 2 つの情報は、単に別装置からのアクセス要求があった場合にポート A ホスト要求情報 4 2 0 を利用し、ディスクアレイ制御部が別装置に対してアクセス要求したい場合に、ポート A 転送指示情報 4 1 0 を利用するにすぎない。したがって、仮にディスクアレイ制御部がホストに対して転送指示をする場合は、ポート A 転送指示情報 4 1 0 を利用することになる。つまり、ホストのポート ID を指定することで、ポート制御部 1 0 0 は、ホストに対して転送する。逆に、バックアップ装置からアクセス要求があった場合は、ポート A ホスト要求情報 4 2 0 を利用する。このように、ポート制御部 1 0 0 は、A 転送指示情報 4 1 0 で指定されたポート ID がホストであったりバックアップ装置であったりと切り替わりながら通信することとなる。

【 0 0 3 1 】

図 5 にポート毎の負荷情報を示す。ポート負荷情報 4 5 0 はポート A に対しての情報であり、ポート負荷情報 4 6 0 はポート B に対しての情報であり、それぞ

れ採取される。本情報は、一定期間内のトータルの I / O 回数を示す I / O 数 4 5 1、転送量 4 5 2、アクセス情報 4 5 3 から構成される。当該ディスクアレイ装置がホストからの I / O 要求を受領し、ディスクアレイ制御部 3 0 0 がこの受領をホスト要求情報から認識した際に、コマンド情報 4 1 2 からデータ転送長とコマンドが取得される。データ転送量に関し、RD (R e a d) の場合は転送長はそのまま、WR (W r i t e) の場合は、4 倍の転送長とする (デュプレックス中でのデータ更新処理に伴うデータ転送回数を考慮して 4 倍としている) 。又、前回のコマンド情報を RAM 4 0 0 内に覚えておいて、今回のコマンドと比較し、連続アクセス (シーケンシャル) か、ランダムアクセスなのかを判断する。そして、I / O 数 4 5 1 をプラスし、転送量 4 5 2 にデータ転送長をプラスし、シーケンシャルかランダムかをアクセス情報 4 5 3 に書き込む。尚、本情報は一定時間毎にクリアされ、単位時間当たりの情報が採取されている。一定時間毎の管理はディスクアレイ制御部 3 0 0 内にあるタイマー機能により実現される。尚、本情報は RAM 4 0 0 に存在する。

【 0 0 3 2 】

図 6 のコントローラ A 負荷情報 6 1 0 のそれぞれは、コントローラ A 内のポート A 負荷情報とコントローラ A 内のポート B 負荷情報の合計した数値が値であり、キャッシュメモリ 5 0 0 の管理情報 6 0 0 内に存在する。本コントローラ A 負荷情報もコマンド受領時にディスクアレイ制御部 3 0 0 が設定する。コントローラ B 負荷情報 6 2 0 も、コントローラ A 負荷情報 6 1 0 と同様である。

【 0 0 3 3 】

図 7 のバックアップ進捗情報 6 3 0 は、キャッシュメモリ 5 0 0 の管理情報 6 0 0 内に存在し、バックアップの対象である論理ボリュームの番号 (対象 LU 番号) 6 3 1、どこまでバックアップが進んでいるかを示すバックアップ実行ポインタ 6 3 2 により構成される。本情報もホストよりバックアップ指示を受領したディスクアレイ制御装置 3 0 0 が、バックアップ対象となった論理ボリュームの番号を 6 3 1 に書き込み、又、バックアップの要求をポート制御部 1 0 0 又は 2 0 0 に行うとともにバックアップ実行ポインタ 6 3 2 の情報を更新する。

【 0 0 3 4 】

コピーポインタ 6 3 3 は、データデュプレックスの処理を実行する際に使用する。コピーポインタ 6 3 3 は、データデュプレックス処理がどこまでデータをデュプレックスしたのかを記録する。データデュプレックス処理はデュプレックスの終わった個所を逐次更新する。これは、バックアップ処理がさらに同時に実行する際、バックアップ処理がコピー実行ポインタ 6 3 3 を参照し、当該ポインタを超えない範囲のデータをバックアップするように制御する。これにより、デュプレックスされた部分を追いかけるようにバックアップ処理をすることが可能である。

【 0 0 3 5 】

図 8 のユーザ指定ポート情報 6 4 0 は、キャッシュメモリ 5 0 0 内の管理情報 6 0 0 に存在し、ユーザーが P C / W S 3 0 0 0 により設定したポートの利用方法を格納しておく。具体的には、当該ポートをバックアップ専用、オンライン専用、またはオンライン／バックアップ共用のいずれかの値を設定する。尚、バックアップ要求があった場合、ディスクアレイ制御部 3 0 0 が本情報を参照し、バックアップポートを検索し、当該ポートのポート制御部に対し、要求を発行する。

【 0 0 3 6 】

図 9 のコピー／バックアップタイミング指示情報 6 5 0 は、キャッシュメモリ 5 0 0 内の管理情報 6 0 0 に存在し、ユーザーが P C / W S 3 0 0 0 により設定した 2 重化の為のコピーとそれに続くバックアップの実行契機を設定する。尚、P C / W S はコントローラに L A N 等で接続されることが多い。

【 0 0 3 7 】

図 1 0 のスケジューリング情報 6 6 0 は、キャッシュメモリ 5 0 0 内の管理情報 6 0 0 に存在し、前記図 9 のコピー／バックアップタイミング指示情報 6 5 0 が時間指定であった場合、当該時間情報を時間監視指示情報 6 6 0 に設定する。

【 0 0 3 8 】

図 1 1 は、優先度情報 6 7 0 のテーブルを示す。これは、ディスクアレイ制御装置がオンライン処理を優先するか、またはバックアップ処理を優先するかを表す値が保持されている。これは、予めユーザが P C / W S 3 0 0 0 を使用して設

定する。バックアップ開始時点でディスクアレイ制御部 3 0 0 は当該情報を参照し、バックアップ優先であれば、あるデータ量までは占有して実行し、当該処理が終了したら、オンライン要求があるかを判断する。なければ、又一定量のデータ量をバックアップ処理する。逆にオンライン処理優先であれば、一定時間はオンライン処理のみを行い、その後、一定量のバックアップデータを転送し、その後又一定時間はオンライン処理を実行する。バックアップの一定量が大きければその分バックアップの優先度は高くなり、オンライン処理に占有する一定時間が長くなればその分だけオンライン処理の優先度は高くなる。以上、コントローラ A を例に説明したがコントローラ B も同様な構成を持つ。

【 0 0 3 9 】

次にディスクアレイ制御装置の基本的な R D (R e a d) / W R (W r i t e) 動作について説明する。ホストからの R D 要求時、ポート制御部 1 0 0 が前記要求を受領した時、ポート A ホスト要求情報 4 2 0 に設定する。ディスクアレイ制御部 3 0 0 は上記情報 4 2 0 を参照し、R D 要求であることを認識し、要求 A D R とデータ長を認識する。当該対象データがキャッシュメモリ 5 0 0 内に存在する場合、当該キャッシュ A D R 情報をポート A 制御部に知らせ、R D 要求したホストへの転送の指示を行う。又、キャッシュメモリ上に存在しない場合は対象となるデータをディスク装置群 4 0 0 0 より読みだし、キャッシュメモリ 6 0 0 に格納し、転送する。

【 0 0 4 0 】

ホストからの W R 要求時、ポート制御部 1 0 0 が前記要求を受領した時、ポート A ホスト要求情報 4 2 0 に設定する。ディスクアレイ制御部 3 0 0 は上記 4 2 0 情報を参照し、W R 要求であることを認識し、要求 A D R とデータ長を認識する。そして、キャッシュメモリ 5 0 0 に転送したところでポート A 制御部に終了報告をするよう伝える。その後、キャッシュメモリ 6 0 0 より、非同期にディスク装置群に掃出し処理（書き込み処理）を行う。

【 0 0 4 1 】

次に本発明の基本機能である、1 ポートでのオンライン処理とバックアップ処理の平行動作について説明する。尚、以下で説明するホスト要求は通常のコマン

ドの一部としてホストから指示されてもいいし、PC/WSよりユーザーが指示してもかまわない

まず、図12において、ポート制御部100を用いてオンライン処理とバックアップ処理を並行して処理するための、通信処理について説明する。なお、図12以降では、ポート制御部がFibreで接続されていることを前提に説明する。

【0042】

Fibre接続の場合、例えばEmulex社のチップセットを利用することが考えら得る。このチップセットの中には2410020 と呼ばれるチップがあり、Fibreプロトコル制御を主に処理することができる。これらのチップセットは、通信処理を多重化することが可能で、例えばホストとの間のオンライン処理とバックアップ装置との間のバックアップ処理の2つを、時分割により同時処理が可能となる。このチップセットを利用した通信多重処理を図12以降のフローチャートで説明する。

【0043】

図12は、イニシエータタスクとターゲットタスクの生成を制御するフローチャートである。ステップ1101では、ディスクアレイ制御装置300からイニシエータタスクの生成要求があるかどうかをポートA制御部100がポートA転送指示情報410を参照することで確認する。例えば、バックアップ処理を実施する場合、ディスクアレイ制御装置300は、イニシエータタスクの生成を要求するためにポート転送指示情報410に転送内容を設定する。ポートA制御部100がこの転送要求を認識した場合、次のステップ1102を実行する。

【0044】

ステップ1102は、要求されたデータ転送を実施するためのイニシエータタスクを生成する。ここでイニシエータタスクと呼んでいるのは、あくまでもポートA転送指示情報で指定されている処理を実行する、という意味で使用する。したがって、本タスクは実際にはデータの送信だけでなく受信も行う。バックアップ装置へのバックアップ処理の場合、送信先のバックアップ装置の情報をパラメータとしてイニシエータタスクを生成する。

【0045】

一方、ステップ 1 1 0 2 で転送要求がなければ、他のホストからの接続要求があるかどうかを確認するステップ 1 1 0 3 を実行する。ステップ 1 1 0 3 は、ポート A 制御部 1 0 0 が他の記憶装置またはホストから転送要求があるかどうかをチェックする。転送要求とは、例えば、相手装置からの Login 要求である。相手装置からの転送要求があれば、ステップ 1 1 0 4 を実施する。ステップ 1 1 0 4 は、転送要求に対する転送処理を実施するターゲットタスクを生成する。ここでいうターゲットタスクとは、他の装置からのアクセスに基づいて処理を実行する、という意味で使用する。したがって、本タスクは実際にはデータの受信だけでなく送信も行う。このフローチャートでは、イニシエータタスクとターゲットタスクが各々生成された時点で、ステップ 1 1 0 1 へ戻る。したがって、生成されたタスク処理の完了を待たず、次の転送要求を待つことになる。以上より、イニシエータタスクとターゲットタスクとは、時分割ではあるがほぼ同時に実行する場合がある。このとき、接続先である Fibre channel は、時分割でパケットデータを送出することができるので、ポート A 制御部 1 0 0 が同時に 2 つ以上のタスクを処理しても Fibre channel ネットワークでのデータ送受信が可能となる。

【 0 0 4 6 】

図 1 3 は、イニシエータタスクの処理を表したフローチャートである。イニシエータタスクは、たとえばバックアップ処理のためディスクアレイ制御装置 3 0 0 が他の記憶装置へデータ転送するための通信タスクである。具体的には、ステップ 1 2 0 1 で相手の装置へ接続要求である Login コマンドを発行する。ステップ 1 2 0 2 は Login が認証された場合、実のデータを相手方へ送信する処理である。データ転送では主に送信を行うが、相互のデータ転送のための認証処理等により送受信処理を行っている。データ転送の完了後、ステップ 1 2 0 3 にて Logout を発行する。

【 0 0 4 7 】

図 1 4 は、ターゲットタスクの処理を表したフローチャートである。ターゲットタスクとは、例えばホストからのオンライン処理のように、ホストからのデータ I/O を送受信する場合の処理が該当する。この場合、ホストからの READ コ

マンドのように、コマンドを受信した後にディスクアレイ装置 3 0 0 がホストへデータ送出することが考えられる。

【 0 0 4 8 】

ステップ 1 3 0 1 は、相手装置からの Login に対して認証を行う。認証とは、正しい相手かどうかをチェックすることである。正しい相手かどうかはキャッシュメモリ 5 0 0 内の管理情報 6 0 0 を参照する。この管理情報 6 0 0 に、接続許可するホストの識別子、または接続許可する他外部記憶装置の識別子が保持されている。この情報を参照して決定する。正しい相手であると認証した場合、次のステップ 1 3 0 2 を実行するが、不正な相手であればエラー応答し、処理を中断する。ステップ 1 2 0 2 は、相手からのデータ転送要求を処理する。データ転送では主に受信を行うが、相互のデータ転送のための認証処理等により送受信処理を行っている。データ転送完了後、ステップ 1 2 0 3 にて Logout を発行する。ステップ 1 3 0 3 は、相手装置とのデータ転送終了後、相手装置から Logout が発行されるので、Logout を検知した場合、それを認証するステップである。

【 0 0 4 9 】

以上のフローチャートにより、イニシエータタスクとターゲットタスクを多重処理でき、1ポートであってもオンライン処理およびバックアップ処理を同時に処理することが可能となる。

【 0 0 5 0 】

次に、ポートが 2 つ以上あった場合について言及する。
この場合は、ポートが 2 つ以上存在するので、オンライン処理とバックアップ処理を行うポートを各々 1 つのポートに割り付けることができる。その際、各ポートの付加状況に応じてどのポートを選択すべきかを判断する必要がある。以下に、2 つのコントローラ A, B から構成するディスクアレイ装置 5 0 0 0 を用いてその論理を説明する。

【 0 0 5 1 】

まず、ディスクアレイ制御部 3 0 0 は、コントローラ A 負荷情報 6 1 0 およびコントローラ B 負荷情報 6 2 0 の各々の負荷情報 6 1 0, 6 2 0 に含まれる転送量 6 1 2 を比較する。転送量の大きい方を負荷が高いコントローラとして判断す

る。もし単位時間当たりの転送量が両方ともあまり高くない場合は I / O 数 6 1 1 で負荷を判断する。ここで仮に、コントローラ A の方が負荷が高いと判断された場合、ディスクアレイ制御部 3 0 0 は、コントローラ A 内の R A M 4 0 0 に格納されているポート負荷情報を参照する。

【 0 0 5 2 】

具体的には、ポート A 負荷情報 4 5 0 内のアクセス情報 4 5 3 を参照する。アクセス情報 4 5 3 は、ホストからの I / O がシーケンシャルかランダムかのどちらかを示している。ホストからの I / O がシーケンシャルであった場合当該ポートの負荷が高いと判断する。両方シーケンシャルであった場合、転送量 4 5 2 を比較して負荷を判断する。両方ランダムであった場合も同様、転送量 4 5 2 で比較する。このようにして、負荷の少ないポートを選択することで、他のオンライン処理等への影響を抑えることができる。

【 0 0 5 3 】

次に、バックアップ処理がシーケンシャル処理である点を考慮したポートの選択論理について説明する。ここでは、1 コントローラあたり 2 ポート構成かつ 2 つのコントローラが実装されているディスクアレイ装置 5 0 0 0 の場合について説明する。この場合、ポートは全部で 4 つ実装されていることになる。

【 0 0 5 4 】

ディスクアレイ装置 5 0 0 0 において、3 つのポートでシーケンシャル処理、残りの 1 つのポートでランダム処理が行われていた場合を想定する。バックアップ処理はシーケンシャル処理であるため、同じシーケンシャル処理を実施しているポートを選ぶと、負荷が余計にかかってしまう。そこで、シーケンシャルに使用しているポートではなく、ランダムアクセスであるポートをバックアップ処理用に使用する。これにより、バックアップ処理も効率よく実施でき、またオンライン処理にも影響を与えないですむ。

【 0 0 5 5 】

次に、ポートの負荷状況に応じたバックアップ処理の引継ぎ論理について説明する。ポートの負荷は逐次変化するので、ポートの負荷状況に応じて、ある時点でのポートの負荷状況を把握して、バックアップ処理を実行していたポートより

も負荷の低いポートを探索し、当該ポートに対してバックアップ処理を引き継がせるのである。ポートの選択方法は次のとおりである。一定時間単位に上記負荷情報を参照し以前選択したポート以外に負荷の低いポートを選択する。バックアップ処理の引継ぎ方法は、次のとおりである。引き継ぐ前のポートは、バックアップ実行中にバックアップ進捗情報 6 3 0 を更新する。バックアップ処理を別のポートに引き継がせた場合、当該ポートはバックアップ進捗情報 6 3 0 を参照し、対象 LU 番号 6 3 1 と実行ポインタ 6 3 2 を得ることで、前のポートが行ってきたバックアップの進捗状況を引き継ぐ。又、夜間等複数のポートの負荷が低いと判断した場合、複数のポートでバックアップ処理を分散して行うことも可能である。具体的には、キャッシュ内にあるバックアップ進捗情報を参照、更新し処理を行う。このようにポートの引継ぎ機能を応用することで、バックアップ処理の効率化を図ることができ、夜間集中バックアップ処理、土日のバックアップ集中処理等、ユーザ要求の高いバックアップ処理を実現することが可能となる。

【 0 0 5 6 】

また、バックアップの開始指示は、ホストからの指示だけではなく、P C / W S から指示することも可能である。P C / W S は、バックアップ処理の起動指示をディスクアレイ制御部 3 0 0 に対して行う。このとき、ユーザがバックアップ処理を実行するポートを直接指定することも可能である。また、P C / W S からポート毎に、ポートに割り付ける処理の種類を指定することも可能である。例えばオンライン処理専用、バックアップ専用、オンライン／バックアップ両用等である。本情報をディスクアレイ制御部が参照し、バックアップ処理を行うポートを選択する。さらに、ユーザはオンライン処理を優先してバックアップ処理を行うよう指示することもできる。又、劣化の割合も、ユーザが指定することが可能である。

【 0 0 5 7 】

さらに、バックアップ指示をユーザが行うのではなく、予めバックアップ開始時刻を設定し、設定された時刻に P C / W S が自動的にバックアップ指示を行うことも可能である。

【 0 0 5 8 】

以上のように、ポートの選択処理、バックアップ処理の引継ぎ処理およびユーザからのバックアップ操作について説明した。次に、バックアップ処理が障害で中断した場合、またはオンライン処理とは別のボリュームコピー処理を実行させた場合の、バックアップ処理について説明する。

【 0 0 5 9 】

まず、バックアップ処理中障害が発生した場合について説明する。
コントローラ障害時、障害制御部 9 0 0 が他系コントローラの障害を認識し、キャッシュメモリ 5 0 0 内のバックアップ進捗情報 6 3 0 を参照し、バックアップ処理を継続する。これにより、障害の発生したコントローラをディスクアレイ装置より切り離すことができるので、障害発生個所のコントローラを交換して障害回復させることができる。障害回復後、先に説明したバックアップ処理の棺処理を行うことも可能となる。

【 0 0 6 0 】

次に、ボリュームコピー処理との同時実行について説明する。ボリュームコピー処理とは、対象となる論理ボリュームデータをコピーしてデュプレックスする処理である。任意の論理ボリュームに対してボリュームコピー処理を実行すると、正（コピー元）と副（コピー先）の 2 つの論理ボリュームが物理ディスク装置群内に存在することになる。このとき、コピー元の正論理ボリュームに対してホストからの I / O があると、その結果を副側に反映させることで二重状態を維持させる。

【 0 0 6 1 】

この機能はバックアップ処理において次のように利用される。つまり、オンライン処理をコピー元である正論理ボリュームに対して実施し、副論理ボリュームに対してバックアップ処理を行う。バックアップ処理中、二重状態を一度中断させることで、中断時点での副論理ボリュームに記録されているデータをバックアップすることが可能となる。

【 0 0 6 2 】

このとき、正論理ボリューム容量が増大するとデュプレックス処理が高負荷となりオンライン処理に多大な影響を与えてしまう場合がある。また、デュプレッ

クス処理の終了後にバックアップ処理が行われる為、バックアップ処理が長時間かかってしまう。そこで、本発明ではボリュームコピー処理とバックアップ処理を同時実行することにより、バックアップデータ採集時間の短縮と、サブシステム内の負荷を減少し、パフォーマンスの劣化を防ぐことが可能である。

【 0 0 6 3 】

元来、通常の正論理ボリュームからキャッシュメモリへのステージング、キャッシュメモリから副論理ボリュームへの書き込み、副論理ボリュームからのバックアップデータ転送の為のキャッシュメモリへのステージング、キャッシュメモリからの F i b r e C h a n n e l へのデータ転送、とコントローラ内のキャッシュメモリ 5 0 0 を介した転送路を何度もデータが転送される。これにより、内部バスの負荷が増大し、システムとしてオンライン処理のパフォーマンスが劣化する。

【 0 0 6 4 】

負荷を低減させる為には、正論理ボリュームからデュプレックスのためのデータをキャッシュメモリ 5 0 0 へステージングした時、当該ステージングデータをそのまま、バックアップデータとして転送することで実現できる。これにより、バスの負荷を低減し、オンライン処理のパフォーマンスの劣化を防ぐことができる。論理ボリュームのデュプレックス処理とバックアップ処理を同時に行うことができるので、全体としてのバックアップ処理が短時間で出来るようになる。

【 0 0 6 5 】

これらのボリュームコピー処理とバックアップ処理との組み合わせから様々な利用方法が考えられ、要求・運用も多岐にわたる。例えば、あるアプリケーションは、あるタイミングでの論理ボリュームのデータのレプリカが必要で、かつ、バックアップのタイミングは週末が望ましいといった要求である。本発明では、そのような様々な要求に満たすべくコピーのタイミングとバックアップのタイミングをユーザーが P S / W S を通じて指定できるようにする。デュプレックスの指示があった時、コピー／バックアップタイミング指示情報 6 5 0 の情報をディスクアレイ制御部 3 0 0 が参照し、同時指定されていれば上述方法にて同時制御を行う。同時指定されていなければ、コピー処理を行う。尚、バックアップのタ

イミングで時間指定、例えば、夜間 24 時を過ぎたらバックアップ処理を開始する等、時間監視指示情報 660 に設定することも可能である。本情報はディスクアレイ制御部が一定時間毎に現時刻と参照し、当該指示時間になったら、バックアップ処理を開始することにより実現される。

【0066】

【発明の効果】

本発明によれば、ストレージエリアネットワークに代表されるネットワーク環境にて Fibre Channel にて接続されたディスクアレイ制御装置に於いて、1ポート構成に於いても、オンライン処理とバックアップ処理を同時実行が可能な制御装置を提供可能である。又、負荷に応じたバックアップ処理や指定時間でのバックアップ処理等が可能であり、オンライン処理へのパフォーマンスを維持することが可能である。さらにコントローラ障害、ポート障害時でもバックアップ処理を他のポート、及び他のコントローラが継続することが可能である。

【0067】

コピー処理とバックアップ処理を同時に制御でき、これらの組み合わせをユーザが設定・実行可能とすることができる。

【図面の簡単な説明】

【図1】

本発明の実施例であるネットワーク環境の構成図である。

【図2】

本発明の実施例であるディスクアレイ制御装置の構成図である。

【図3】

本発明の実施例であるディスク装置群の構成図の一例である。

【図4】

本発明の実施例であるポート制御部ディスク制御部間 I/F 情報の一例である。

【図5】

本発明の実施例であるポート負荷情報の一例である。

【図 6】

本発明の実施例であるコントローラ負荷情報の一例である。

【図 7】

本発明の実施例であるバックアップ進捗情報一例である。

【図 8】

本発明の実施例であるユーザ指定ポート情報の一例である。

【図 9】

本発明の実施例であるコピー／バックアップタイミング情報の一例である。

【図 1 0】

本発明の実施例であるスケジューリング情報の一例である。

【図 1 1】

本発明の実施例である優先度情報の一例である。

【図 1 2】

本発明の実施例であるポート制御部 1 0 0 内のイニシエータタスクまたはターゲットタスクを生成するフローチャートである。

【図 1 3】

本発明の実施例であるポート制御部 1 0 0 内のイニシエータタスクのフローチャートである。

【図 1 4】

本発明の実施例であるポート制御部 1 0 0 内のターゲットタスクのフローチャートである。

【符号の説明】

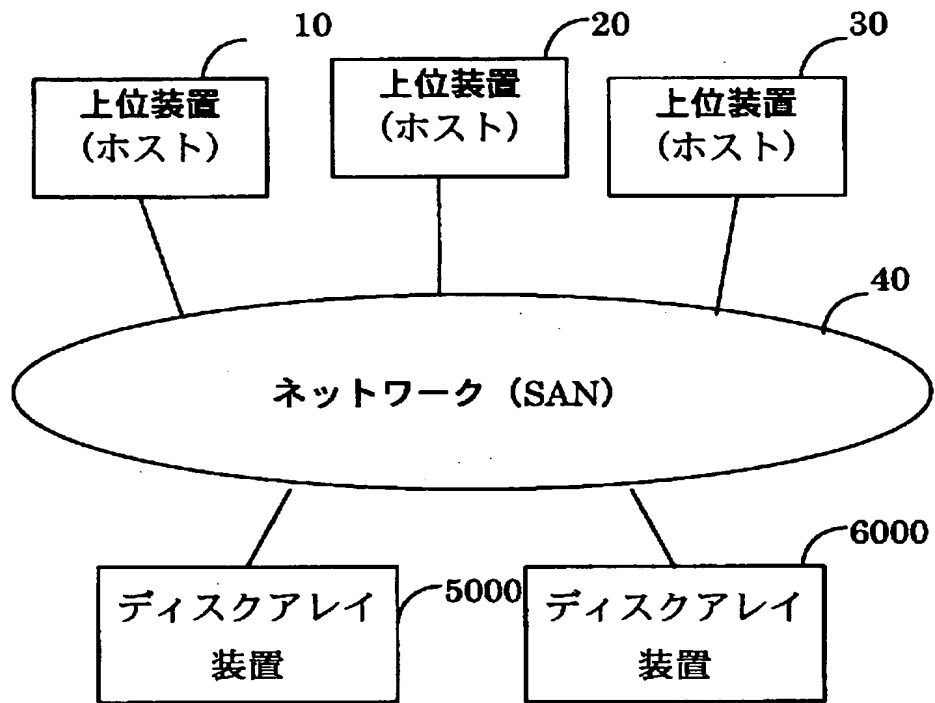
1 0、2 0、3 0 : ホストコンピュータ、4 0 : F i b r e C h a n n e l (F a b r i c S W)、1 0 0、2 0 0 : ポート制御部、3 0 0 : ディスクアレイ制御部、
4 0 0 : R A M、5 0 0 : キャッシュメモリ、6 0 0 : キャッシュメモリ内管理情報、7 0 0、8 0 0 : ドライブ制御部、9 0 0 : 障害制御部、1 0 0 0、2 0 0 0 : コントローラ部、3 0 0 0 : P C / W S、4 0 0 0 : ディスク装置群、5 0 0 0、6 0 0 0 : ディスクアレイ装置、

特 2 0 0 1 - 0 5 7 6 3 1

【書類名】 図面

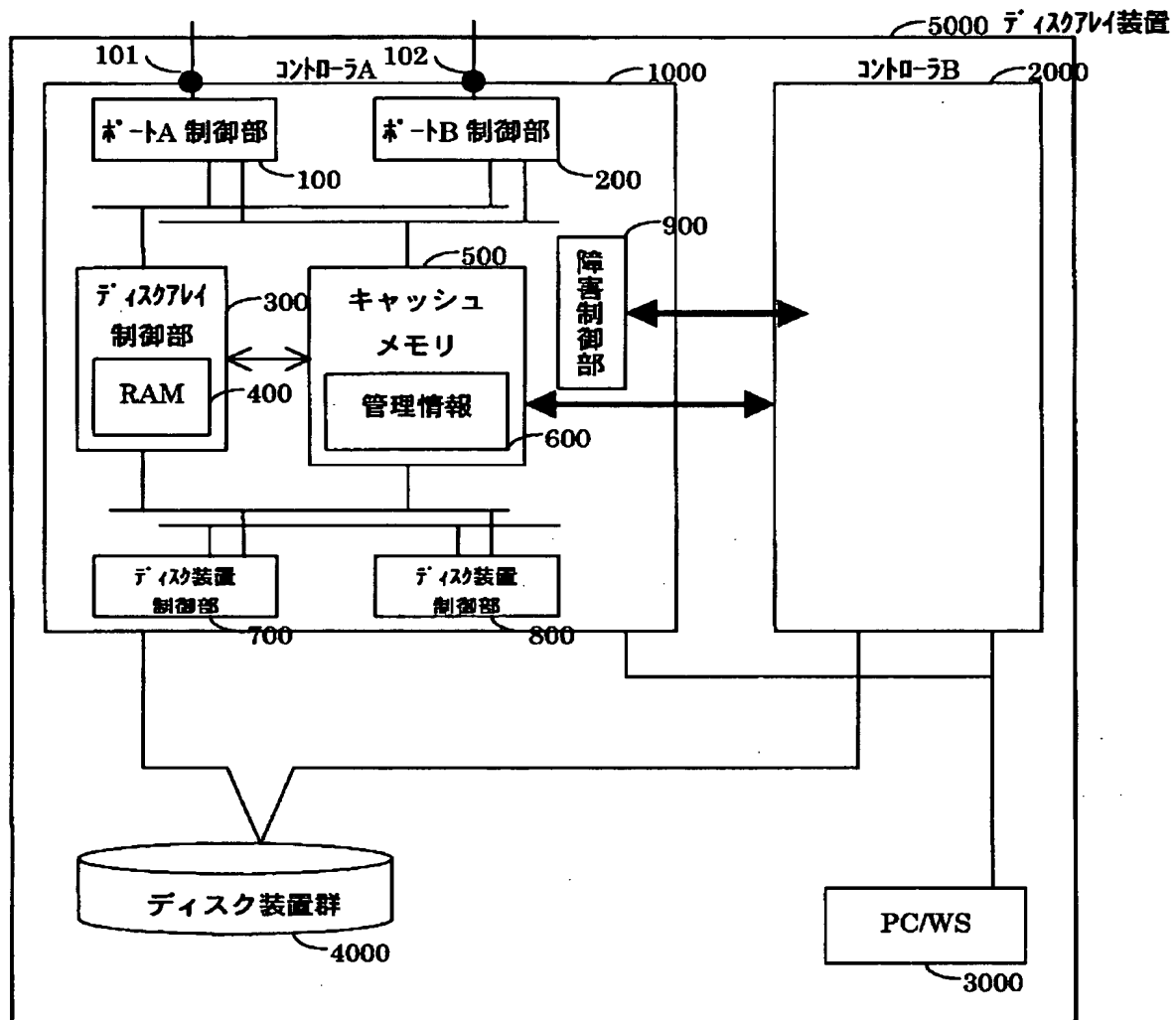
【図 1】

図 1



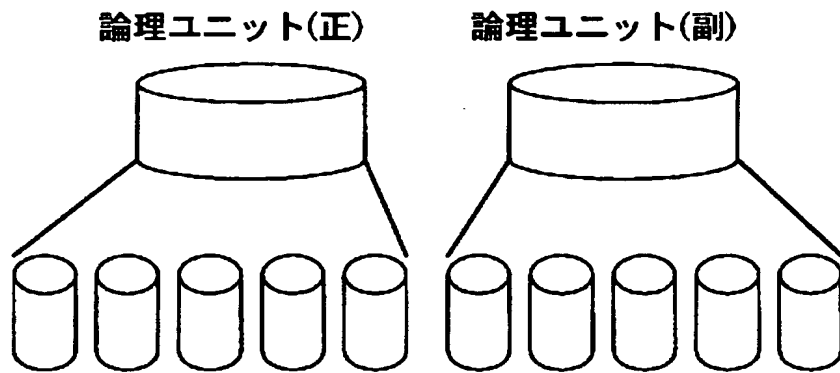
【図 2】

図 2



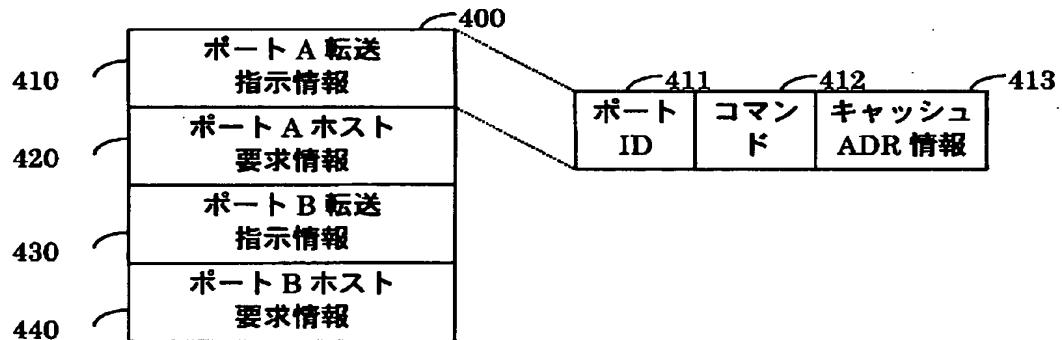
【図 3】

図 3



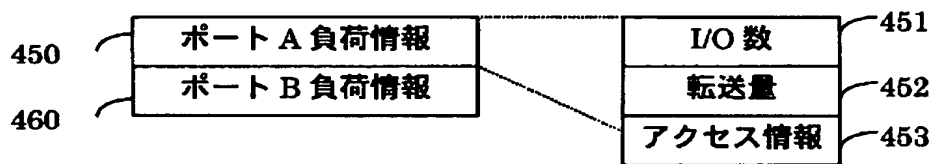
【図 4】

図 4



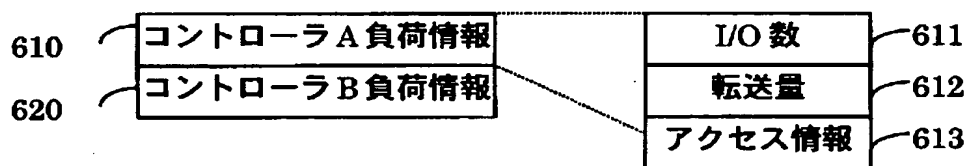
【図 5】

図 5



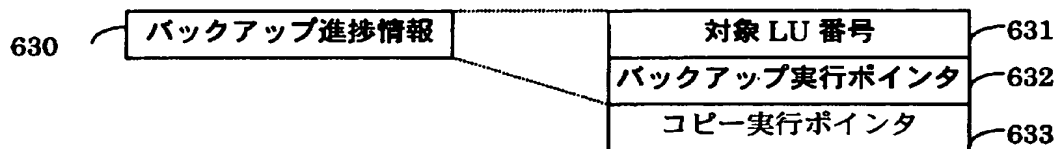
【図 6】

図 6



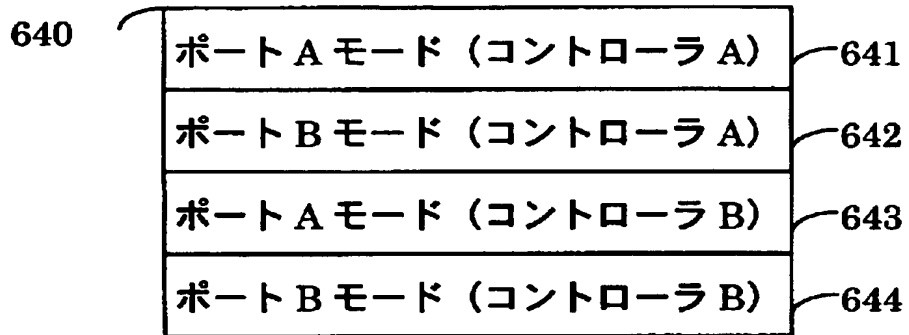
【図 7】

図 7



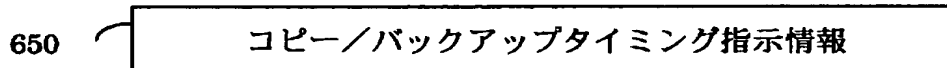
【図 8】

図 8



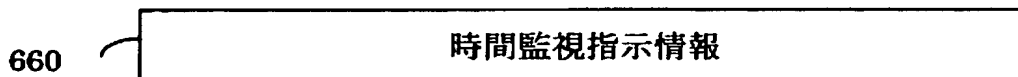
【図 9】

図 9



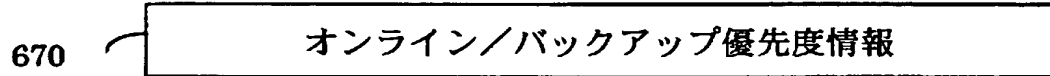
【図 10】

図 10



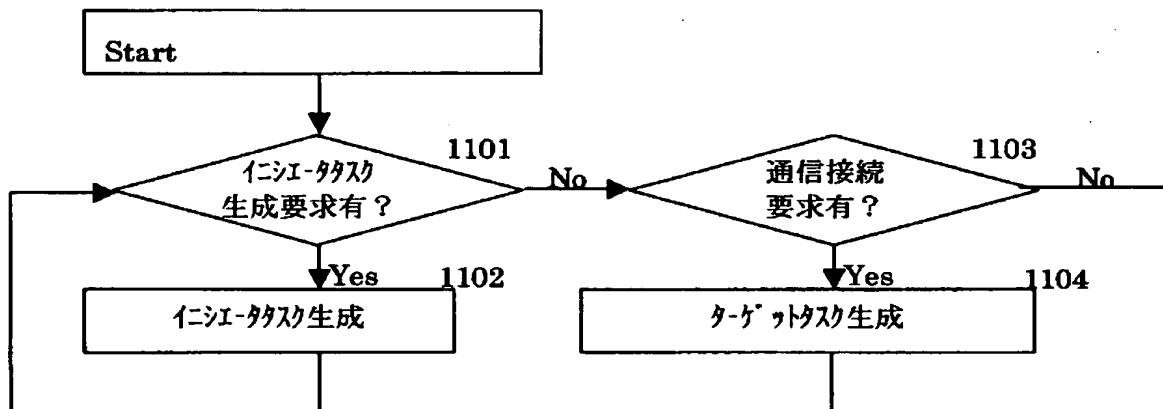
【図 1 1】

図 11



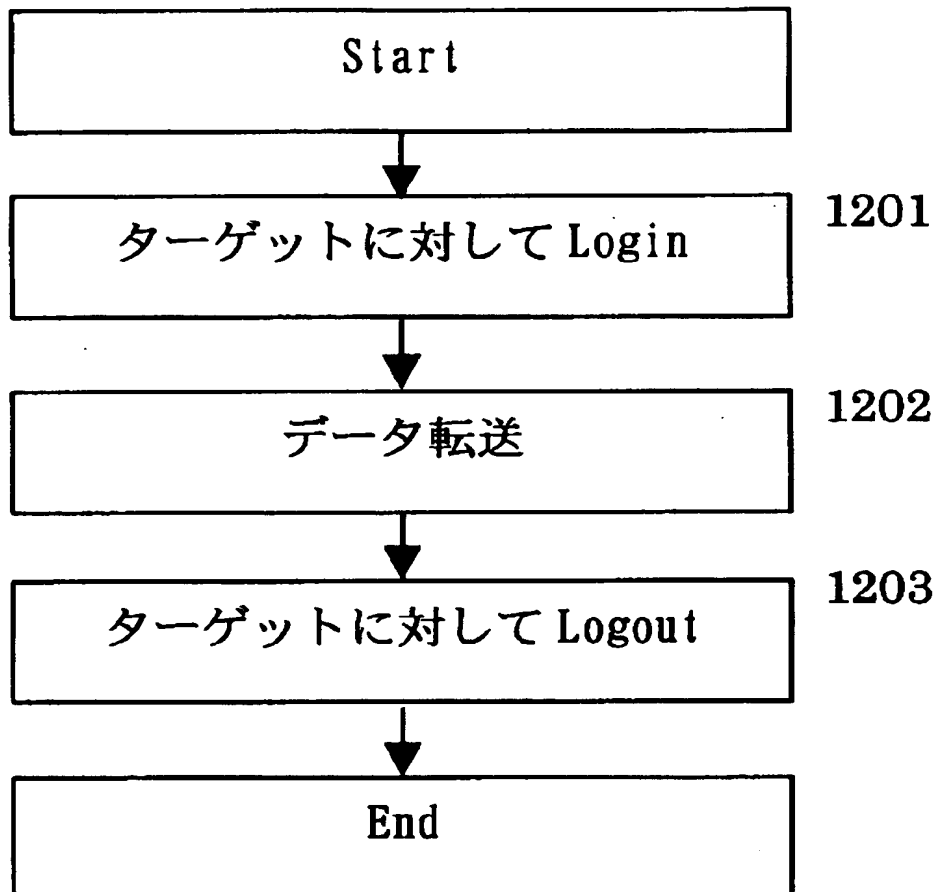
【図 1 2】

図 12



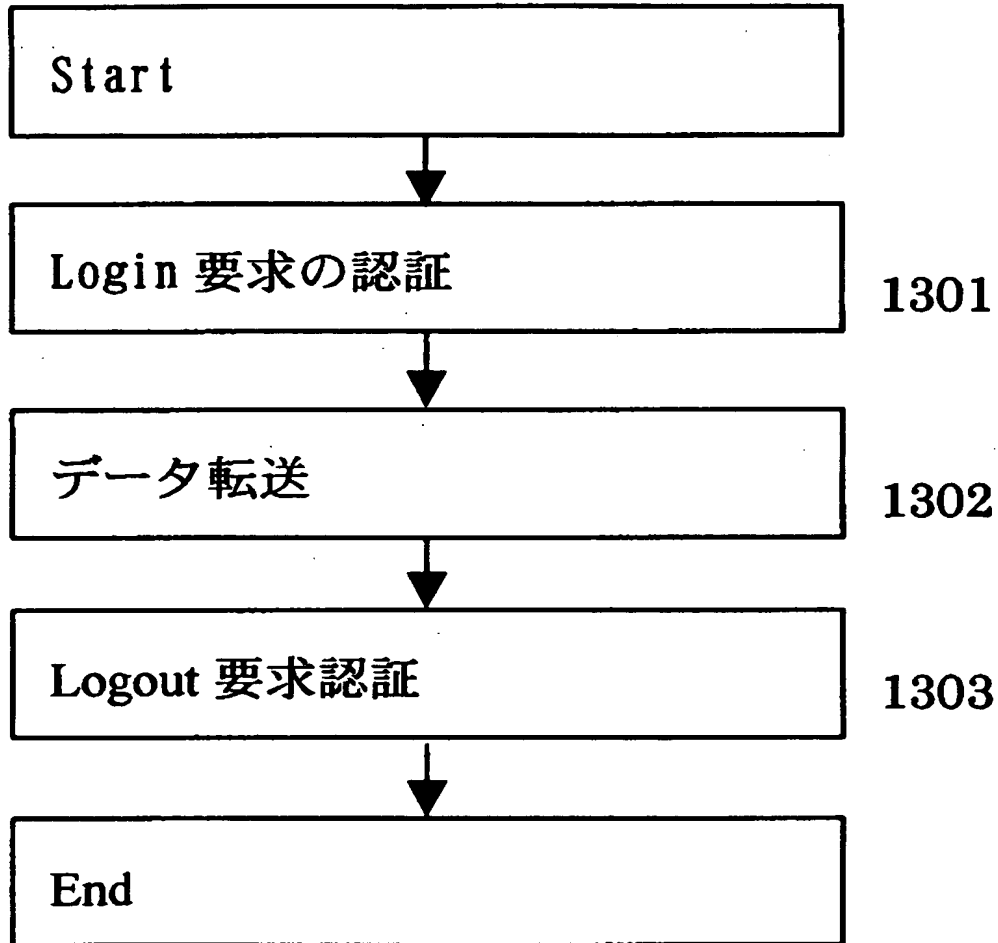
【図 1 3】

図 13



【図 1 4】

図 14



【書類名】 要約書

【要約】

【課題】

本発明の目的は、ストレージエリアネットワーク環境でFibre Channelにて接続されたディスクアレイ装置でコントローラ当たり1プロセッサにて制御される1ポート又は複数ポートを備えている時、1ポート、1プロセッサにおいても、オンライン業務とバックアップ業務をオンラインの負荷を考慮して同時実現可能な装置を提供することにある。

【解決手段】

本発明では、ポート制御部にホストからの要求を受けるだけでなく、他の記憶制御装置に対し、要求を発行できる機能を合わせ持つことによりオンライン／バックアップ同時処理を可能とする。また、複数ポート時は負荷に応じたポート選択やスケジューリングを行う事によりバックアップによるオンライン業務へのパフォーマンスの劣化の割合を低減する。

【効果】

これにより、オープン向けの中小規模ディスクアレイ装置においても、通常業務実行中でのバックアップ業務の実現とユーザの運用に見合ったバックアップ業務の提供とオンライン業務の性能劣化を抑えたバックアップ業務を実現することができる。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所